

Quantum Similarity Measures under Atomic Shell Approximation: First Order Density Fitting Using Elementary Jacobi Rotations

LLUÍS AMAT, RAMON CARBÓ-DORCA

Institute of Computational Chemistry, University of Girona, Girona 17071, Catalonia, Spain

Received 2 May 1997; accepted 30 July 1997

ABSTRACT: The elementary Jacobi rotations technique is proposed as a useful tool to obtain fitted electronic density functions expressed as linear combinations of atomic spherical shells, with the additional constraint that all coefficients are kept positive. Moreover, a Newton algorithm has been implemented to optimize atomic shell exponents, minimizing the quadratic error integral function between *ab initio* and fitted electronic density functions. Although the procedure is completely general, as an application example both techniques have been used to compute a 1*S*-type Gaussian basis for atoms H through Kr, fitted from a 3-21G basis set. Subsequently, molecular electronic densities are modeled in a *promolecular* approximation, as a simple sum of parameterized atomic contributions. This simple molecular approximation has been employed to show, in practice, its usefulness to some computational examples in the field of molecular quantum similarity measures. © 1997 John Wiley & Sons, Inc. *J Comput Chem* **18**: 2023–2039, 1997

Keywords: atomic shell approximation; Carbó Index; elementary Jacobi rotations; promolecular densities; quadratic error integral function; quantum similarity measures

Correspondence to: R. Carbó-Dorca

Contract/grant sponsor: CICYT; contract/grant number: SAF 96-0158

Contract/grant sponsor: Spanish Ministerio de Educación y Cultura

Introduction

The elementary Jacobi rotations (EJR) technique as a source of metric-preserving orthogonal transformations has been used in many fields of quantum chemistry. Besides the task of electronic energy optimization,¹ EJR's have been employed in other areas such as the localization of orbitals² and the variational determination of orbitals in molecular pair wave functions.³ In our laboratory, a large contribution of EJR to direct electronic energy optimization has been developed⁴ as well as some earlier applications related to quantum similarity.⁵ Also, work has been performed on the many aspects of the Jacobi diagonalization algorithm,⁶ proposing a new parallelizable procedure, constituting a practical computational scheme,⁷ able to deal with large matrices and a chosen set of eigenvalues and eigenvectors. Recently, an application of EJR to the study of approximate full CI wave functions has also been reported.⁸

The present article will deal, in a general manner, with the proposal of using the EJR technique to solve the problem of fitting the usual first order electronic density functions to a linear combination of *nS-type* spherical functions, as has been done by several investigators.^{9,10} However, this work demands the fulfillment of the additional fitting constraint, which is that all coefficients have to be positive.¹¹ This is in order to preserve the *statistical meaning* of the density function,¹² which has to be considered as a probability distribution and thus it should be positive definite *everywhere*. The fitted approximate density employed in this work has been modulated using a superposition of several atomic spherical function squared modules, the so-called atomic shell approximation (ASA).¹¹ In this work only atomic density fittings have been performed and subsequent application to molecular quantum similarity measures (QSM) evaluation^{13–15} will be carried out using a *promolecular* approximation.¹⁶ The main advantage of fitting *nS-type* functions for practical application purposes is the low computational costs compared to *ab initio* calculations. In this way, the feasibility of QSAR project studies involving big sets of large sized molecules of biological interest can be envisaged.¹⁷

To achieve this objective, the present study will first study the description of QSM and ASA-type

functions. The constrained least-squares fitting problem, consisting of minimizing the quadratic error integral function between *ab initio* and ASA electronic density functions, will be studied next. Apart from the least-squares EJR technique, a Newton algorithm will also be constructed to find the optimal ASA exponents. The corresponding computational algorithms will be detailed and some illustrative examples discussed. Finally, using the *promolecular* approximation some molecular QSM will be presented as an application example.

Quantum Similarity Measures

A QSM is defined as the integral involving at least two density functions, $\{\rho_A(\mathbf{r}_1), \rho_B(\mathbf{r}_2)\}$, and a given positive definite operator $\Omega(\mathbf{r}_1, \mathbf{r}_2)$, where the vector \mathbf{r} contains the coordinates of the involved electrons. The QSM integral may be formally written as:

$$Z_{AB} = \iint \rho_A(\mathbf{r}_1) \Omega(\mathbf{r}_1, \mathbf{r}_2) \rho_B(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (1)$$

When both compared systems are the same, then the integral is denoted as Z_{AA} , and referred to as a quantum self-similarity measure (QS-SM), and when the quantum systems studied are molecules, the literature speaks of molecular QSM (MQSM). The first MQSM defined, and also the most common, is the so-called overlap type.^{14a} This choice corresponds to the replacement of $\Omega(\mathbf{r}_1, \mathbf{r}_2)$ by a Dirac delta function, transforming eq. (1) into:

$$Z_{AB} = \int \rho_A(\mathbf{r}) \rho_B(\mathbf{r}) d\mathbf{r} \quad (2)$$

Accurate MQSM may be obtained using a LCAO approach to describe molecular electronic density functions, corresponding to the expression:

$$\rho_A = \sum_{\mu, \nu \in A} D_{\mu\nu} \chi_\mu^*(\mathbf{r}) \chi_\nu(\mathbf{r}) \quad (3)$$

where $D_{\mu\nu}$ is the charge-bond order matrix, and $\{\chi_\mu\}$ the atomic orbital basis set functions. Using this approximation, four-center integrals have to be computed in MQSM. Besides these computational difficulties, there is the additional problem of the maximization of MQSM:¹⁸ the function, Z_{AB} , depends on the relative position of molecules *A* and *B*, and the similarity between them is usually defined as the maximum. It is not surprising,

therefore, that the major disadvantage of *ab initio* calculations is the fact that quantum similarity integral computation is cumbersome. For this reason computational algorithms have been developed to calculate approximate density functions.^{9–11}

QSM theory is an ideal mathematical construction, logically placed into the structure of quantum mechanics,^{13b,c} which may be used whenever it is necessary to compare two or more density functions. In fact, this may be seen from a more general point of view, associating the density functions in eq. (1) with square summable definite positive functions. From the mathematical point of view, a similarity integral defined as Z_{AB} , can be interpreted as positive-valued, weighted scalar products.

Atomic Shell Approximation

The construction of ASA-like density functions can be traced up to the initial studies on QSM,^{14a} where a CNDO-like¹⁹ approach was invoked to deal with the computational problem of the evaluation of the quantum similarity integrals, Z_{AB} , for molecules. From this initial viewpoint, the concept of approximating a given density function, has led to the consideration of ASA as a general superposition of spherical *nS-type*, STO, or GTO, functions.

The density function in ASA form may be written in terms of atomic function set contributions $\{\sigma_a\}$:

$$\rho_A^{\text{ASA}}(\mathbf{r}) = \sum_{a \in A} \sigma_a(\mathbf{r}) \quad (4)$$

where the sum runs over all the atoms present in molecule *A*. At the same time, the atomic set of functions $\{\sigma_a\}$, may be constructed using a set of atomic *shells* $\{s_i\}$, using the sum:

$$\sigma_a(\mathbf{r}) = \sum_{i \in a} s_i(\mathbf{r}) \quad (5)$$

and the sum is over all the atomic *shells* of the *a*th atom. Finally, the spherical function set $\{s_i\}$, can be defined as:

$$s_i(\mathbf{r}) = \sum_{k \in i} c_k \varphi_k(\mathbf{r}) \quad (6)$$

where the sum is carried out over the entire positive definite function set $\{\varphi_k\}$, belonging to shell *i*. The coefficients $\{c_k\}$ must be *positive in all cases* to keep the distribution structure of the approximated function, ρ_A^{ASA} , positive definite.

The above ASA partition is equivalent to writing eq. (4) in a more compact notation, as a linear combination of a positive definite function set $\{\theta_i\}$:

$$\rho_A^{\text{ASA}}(\mathbf{r}) = \sum_{i \in A} w_i \theta_i(\mathbf{r}) \quad (7)$$

where the sum is performed over all the basis function set $\{\theta_i\}$, and the set of positive coefficients $\{w_i\}$ must be determined.

The most interesting case is constituted by the ASA approximation of first order density functions, but other high level density function forms may be considered as well. In any case, both the exact and the ASA density functions may be supposed to be normalized to one particle, by dividing by the appropriate particle number factors, and in eq. (7), considering the basis functions normalized as follows:

$$\int \theta_i(\mathbf{r}) d\mathbf{r} = 1, \forall i \in A \quad (8)$$

then, necessarily the set of ASA coefficients $\{w_i\}$, apart from the condition:

$$w_i > 0, \forall i \in A \quad (9)$$

must fulfill the additional constraint:

$$\sum_{i \in A} w_i = 1 \quad (10)$$

Although the second condition may be easily taken into account with a Lagrange multiplier technique,²⁰ the first one, as expressed in eq. (9), could not be so easily introduced into the computational algorithm.¹¹ It will be shown here that both conditions can be maintained through the optimization process by using the appropriate tools.

PROMOLECULAR ATOMIC SHELL APPROXIMATION

Recently, similarity measures applied to QSAR studies²¹ have gained more importance in computational chemistry. When QSMs are used in QSAR studies, ASA will represent a considerable saving in computation time. However, this approach is still insufficient when a vast number of large molecules are studied. This is due to the necessity of obtaining both the *ab initio* electronic density and the fitted functions (optimal coefficients and exponents) for each molecule.¹¹ A further theoretical and computational simplification is achieved

when the so-called *promolecular* approximation is considered, and then QSAR studies can be performed¹⁷ without much effort. Within this approach, the total molecular electronic density is calculated as a sum of atomic electronic density contributions:

$$\rho_A^{\text{ASA}}(\mathbf{r}) = N_A^{-1} \sum_{a \in A} N_a \rho_a^{\text{ASA}}(\mathbf{r}) \quad (11)$$

where N_A is the total number of electrons in molecule A , and N_a the atomic number of each atom a . The sum in eq. (11) runs over all the molecular atoms.

Every atomic density function ρ_a^{ASA} is constructed with the same structure as the one given in eq. (7), but replacing θ_i by a squared *nS-type* spherical function centered on the a th atom and keeping the coefficient convex constraints, given in eq. (9) and (10):

$$\rho_a^{\text{ASA}}(\mathbf{r}) = \sum_{i \in a} w_i |S_i(\mathbf{r} - \mathbf{r}_a)|^2 \quad (12)$$

Using this approximation, overlap-like QSM between two atoms can be expressed by:

$$Z_{ab} = \sum_{i \in a} w_i \sum_{j \in b} w_j Z_{ij} \quad (13)$$

where the Z_{ij} elements belong to the similarity matrix, \mathbf{Z} , and are defined by the integral:

$$Z_{ij} = \int |S_i(\mathbf{r} - \mathbf{r}_a)|^2 |S_j(\mathbf{r} - \mathbf{r}_b)|^2 d\mathbf{r} \quad (14)$$

the similarity matrix being positive definite.

Constrained Fitting of ASA Coefficients

The set of $\{w_i\}$ optimal coefficients, appearing in eq. (12), have been calculated minimizing the quadratic error integral function between *ab initio* and ASA electronic density functions. The usual form of the quadratic error function is:

$$\begin{aligned} \mathcal{E}^{(2)} &= \int |\rho_a(\mathbf{r}) - \rho_a^{\text{ASA}}(\mathbf{r})|^2 d\mathbf{r} \\ &= Z_{aa} + \sum_{i, j \in a} w_i w_j Z_{ij} - 2 \sum_{i \in a} w_i \\ &\quad \times \sum_{\mu, \nu \in a} D_{\mu\nu} \int |S_i(\mathbf{r})|^2 \chi_\mu^*(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r} \quad (15) \end{aligned}$$

where Z_{aa} corresponds to the *ab initio* QS-SM of atom a :

$$Z_{aa} = \int |\rho_a(\mathbf{r})|^2 d\mathbf{r} \quad (16)$$

which is computed within the LCAO approximation, replacing electronic density, $\rho_a(\mathbf{r})$, by the corresponding expression described in eq. (3).

Eq. (15) may be written in matrix form as:

$$\mathcal{E}^{(2)} = Z_{aa} + \mathbf{w}^T \mathbf{Z} \mathbf{w} - 2 \mathbf{b}^T \mathbf{w} \quad (17)$$

where the elements of the vector $\mathbf{b} = \{b_i\}$ are given by the integral:

$$b_i = \sum_{\mu, \nu} D_{\mu\nu} \int |S_i(\mathbf{r})|^2 \chi_\mu^*(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r} \quad (18)$$

and \mathbf{w} is a normalized column vector ($\mathbf{w}^T \mathbf{w} = 1$) containing the ASA coefficients.

The set of positive coefficients $\{w_i\}$ can be easily defined as:

$$w_i = |x_i|^2 \quad (19)$$

fulfilling condition (9). If they are taken as real, then the quadratic error integral function is simply expressed in terms of the squared coefficients as:

$$\mathcal{E}^{(2)} = Z_{aa} + \sum_{i, j \in a} x_i^2 x_j^2 Z_{ij} - 2 \sum_{i \in a} x_i^2 b_i \quad (20)$$

On the other hand, due to the fact that the similarity matrix, \mathbf{Z} , is positive definite, a unitary matrix, \mathbf{U} , can be found such as:

$$\mathbf{U}^+ \mathbf{Z} \mathbf{U} = \mathbf{D} \quad (21)$$

where \mathbf{D} is a diagonal matrix. The first step in the procedure consists in diagonalizing the similarity matrix \mathbf{Z} , to obtain their eigenvalues and eigenvectors. Then, $\{x_i\}$ coefficients are extracted from the most promising normalized eigenvector of the \mathbf{Z} matrix; that is, the one producing the minimal quadratic error. Consequently, the required constraint, specified in eq. (10), is fulfilled from the beginning. Starting from this initial vector, and applying orthogonal EJR over its components, the required conditions in eq. (9) and (10) will be maintained throughout the iterative procedure.

VARIATION OF QUADRATIC ERROR INTEGRAL FUNCTION APPLYING ELEMENTARY JACOBI ROTATIONS

EJRs are good tools to obtain elementary orthogonal transformations usable over vectors or matri-

ces. The origin of such transformation matrices can be found in Jacobi's study,⁶ dating back to the last century. Being orthogonal, EJR may also be viewed as rotation matrices over n -dimensional spaces. Applied to a given n -dimensional vector, an EJR, which is written as $J_{pq}(\alpha)$, transforms the corresponding p and q components, keeping the rest of them invariable. Calling $\mathbf{x} = \{x_i\}$ the vector to be transformed by means of the EJR $J_{pq}(\alpha)$, the transformation is defined by the equations:

$$\begin{aligned}\dot{x}_p &\leftarrow c x_p - s x_q \\ \dot{x}_q &\leftarrow s x_p + c x_q\end{aligned}\quad (22)$$

where c and s are the cosine and sine of the rotation angle α . The norm of the new vector, $\dot{\mathbf{x}} = \{\dot{x}_i\}$, obtained after applying the EJR $J_{pq}(\alpha)$, remains invariable with respect to the initial one.

ASA coefficients are obtained by an optimization procedure, which minimizes $\varepsilon^{(2)}$. Isolating the p and q elements of the vector \mathbf{x} from the rest, eq. (20) gives:

$$\begin{aligned}\varepsilon^{(2)} &= Z_{aa} + x_p^4 Z_{pp} + x_q^4 Z_{qq} + 2 x_p^2 x_q^2 Z_{pq} \\ &+ 2 x_p^2 \sum_{i \neq p, q} x_i^2 Z_{pi} + 2 x_q^2 \sum_{i \neq p, q} x_i^2 Z_{iq} \\ &+ \sum_{i \neq p, q} \sum_{j \neq p, q} x_i^2 x_j^2 Z_{ij} - 2 x_p^2 b_p - 2 x_q^2 b_q \\ &- 2 \sum_{i \neq p, q} x_i^2 b_i\end{aligned}\quad (23)$$

Over this equation, it is easy to apply the EJR $J_{pq}(\alpha)$, thus the variation of $\varepsilon^{(2)}$ respects the active pair of elements $\{p, q\}$, and may be expressed as:

$$\begin{aligned}\delta \varepsilon^{(2)} &= \delta x_p^4 Z_{pp} + \delta x_q^4 Z_{qq} + 2 \delta (x_p^2 x_q^2) Z_{pq} \\ &+ 2 \delta x_p^2 \sum_{i \neq p, q} x_i^2 Z_{pi} + 2 \delta x_q^2 \sum_{i \neq p, q} x_i^2 Z_{iq} \\ &- 2 \delta x_p^2 b_p - 2 \delta x_q^2 b_q\end{aligned}\quad (24)$$

The squared transformed coefficient elements, \dot{x}_p^2 and \dot{x}_q^2 , necessary to obtain expressions for the variations δx_p^2 and δx_q^2 , are easily obtained from eq. (22), as:

$$\begin{aligned}\dot{x}_p^2 &= (c x_p - s x_q)^2 = c^2 x_p^2 + s^2 x_q^2 - 2cs x_p x_q \\ &= x_p^2 + s^2 (x_q^2 - x_p^2) - 2cs x_p x_q \\ \dot{x}_q^2 &= (s x_p + c x_q)^2 = s^2 x_p^2 + c^2 x_q^2 + 2cs x_p x_q \\ &= x_q^2 + s^2 (x_p^2 - x_q^2) + 2cs x_p x_q\end{aligned}\quad (25)$$

and their second-order variations are given by:

$$\begin{aligned}\delta x_p^2 &= (\dot{x}_p^2 - x_p^2) \\ &= -[s^2 (x_p^2 - x_q^2) + 2cs x_p x_q] = -\Delta \\ \delta x_q^2 &= (\dot{x}_q^2 - x_q^2) \\ &= s^2 (x_p^2 - x_q^2) + 2cs x_p x_q = \Delta\end{aligned}\quad (26)$$

The fourth power of the transformed coefficients is defined from the quadratic ones as follows:

$$\begin{aligned}\dot{x}_p^4 &= (\dot{x}_p^2)^2 = (x_p^2 - \Delta)^2 = x_p^4 + \Delta^2 - 2 x_p^2 \Delta \\ \dot{x}_q^4 &= (\dot{x}_q^2)^2 = (x_q^2 + \Delta)^2 = x_q^4 + \Delta^2 + 2 x_q^2 \Delta \\ \dot{x}_p^2 \dot{x}_q^2 &= (x_p^2 - \Delta)(x_q^2 + \Delta) \\ &= x_p^2 x_q^2 + (x_p^2 - x_q^2) \Delta - \Delta^2\end{aligned}\quad (27)$$

and their corresponding variation, δx_p^4 , δx_q^4 , and $\delta(x_p^2 x_q^2)$, are defined by:

$$\begin{aligned}\delta x_p^4 &= (\dot{x}_p^4 - x_p^4) = \Delta^2 - 2 x_p^2 \Delta \\ \delta x_q^4 &= (\dot{x}_q^4 - x_q^4) = \Delta^2 + 2 x_q^2 \Delta \\ \delta(x_p^2 x_q^2) &= (\dot{x}_p^2 \dot{x}_q^2 - x_p^2 x_q^2) = (x_p^2 - x_q^2) \Delta - \Delta^2\end{aligned}\quad (28)$$

Developing the δx_p^4 and δx_q^4 expressions in eq. (28) gives:

$$\begin{aligned}\delta x_p^4 &= s^4 (x_p^2 - x_q^2)^2 + 4c^2 s^2 x_p^2 x_q^2 \\ &+ 4cs^3 (x_p^2 - x_q^2) x_p x_q \\ &- 2 x_p^2 [s^2 (x_p^2 - x_q^2) + 2cs x_p x_q] \\ &= s^4 [(x_p^2 - x_q^2)^2 - 4 x_p^2 x_q^2] \\ &+ 4cs^3 (x_p^2 - x_q^2) x_p x_q \\ &+ 2s^2 x_p^2 [- (x_p^2 - x_q^2) + 2 x_q^2] \\ &- 4cs x_p^3 x_q\end{aligned}\quad (29)$$

and:

$$\begin{aligned}\delta x_q^4 &= s^4 (x_p^2 - x_q^2)^2 + 4c^2 s^2 x_p^2 x_q^2 \\ &+ 4cs^3 (x_p^2 - x_q^2) x_p x_q \\ &+ 2 x_q^2 [s^2 (x_p^2 - x_q^2) + 2cs x_p x_q]\end{aligned}$$

$$\begin{aligned}
 &= s^4 \left[(x_p^2 - x_q^2)^2 - 4x_p^2 x_q^2 \right] \\
 &\quad + 4cs^3 (x_p^2 - x_q^2) x_p x_q \\
 &\quad + 2s^2 x_q^2 \left[(x_p^2 - x_q^2) + 2x_p^2 \right] + 4cs x_p x_q^3
 \end{aligned} \quad (30)$$

whereas the term $\delta(x_p^2 x_q^2)$ adopts the form:

$$\begin{aligned}
 \delta(x_p^2 x_q^2) &= s^2 (x_p^2 - x_q^2)^2 + 2cs (x_p^2 - x_q^2) x_p x_q \\
 &\quad - s^4 (x_p^2 - x_q^2)^2 - 4c^2 s^2 x_p^2 x_q^2 \\
 &\quad - 4cs^3 (x_p^2 - x_q^2) x_p x_q \\
 &= -s^4 \left[(x_p^2 - x_q^2)^2 - 4x_p^2 x_q^2 \right] \\
 &\quad - 4cs^3 (x_p^2 - x_q^2) x_p x_q \\
 &\quad + s^2 \left[(x_p^2 - x_q^2)^2 - 4x_p^2 x_q^2 \right] \\
 &\quad + 2cs (x_p^2 - x_q^2) x_p x_q
 \end{aligned} \quad (31)$$

Substituting expressions δx_p^2 , δx_q^2 , δx_p^4 , δx_q^4 , and $\delta(x_p^2 x_q^2)$ into eq. (24), and collecting terms, one has:

$$\delta \mathcal{E}^{(2)} = E_{04} s^4 + E_{13} cs^3 + E_{02} s^2 + E_{11} cs \quad (32)$$

where the new parameter set $\{E_{cs}\}$ introduced in eq. (32) is defined using:

$$\begin{aligned}
 E_{04} &= (Z_{pp} + Z_{qq} - 2Z_{pq}) \left[(x_p^2 - x_q^2)^2 - 4x_p^2 x_q^2 \right] \\
 E_{13} &= 4(Z_{pp} + Z_{qq} - 2Z_{pq}) (x_p^2 - x_q^2) x_p x_q \\
 E_{02} &= 4(Z_{pp} + Z_{qq} - 2Z_{pq}) x_p^2 x_q^2 - 2(x_p^2 - x_q^2) G \\
 E_{11} &= -4x_p x_q G
 \end{aligned} \quad (33)$$

and

$$\begin{aligned}
 G &= \sum_{i \neq p, q} x_i^2 (Z_{pi} - Z_{qi}) + x_p^2 Z_{pp} - x_q^2 Z_{qq} \\
 &\quad - (x_p^2 - x_q^2) Z_{pq} - b_p + b_q
 \end{aligned}$$

The optimal sine belonging to the EJRC can be chosen with the gradient condition $\frac{d\delta \mathcal{E}^{(2)}}{ds} = 0$; that is:

$$\begin{aligned}
 \frac{d\delta \mathcal{E}^{(2)}}{ds} &= 4E_{04} s^3 + E_{13} (-ts^3 + 3cs^2) \\
 &\quad + 2E_{02} s + E_{11} (-ts + c) \\
 &= -c[(E_{13} s^2 + E_{11})t^2
 \end{aligned}$$

$$\begin{aligned}
 &-2(2E_{04} s^2 + E_{02})t - (3E_{13} s^2 + E_{11})] \\
 &= -c(T_1 t^2 - 2T_2 t - T_3) = 0
 \end{aligned} \quad (34)$$

where $t = \frac{s}{c}$ and $\frac{dc}{ds} = -t$.

The best EJRC angle is found by solving the second-degree equation appearing in expression (34). The optimization is conducted by an iterative procedure, until the variation of the EJRC angle or the quadratic error integral function value becomes negligible.

Optimization of ASA Exponents

A Newton algorithm²⁰ has been used to optimize the exponents of the fitted *nS-type* functions in ASA. This method is operative when the analytic gradient vector and the Hessian matrix are available, as in the present ASA case. Newton's equation may be written as:

$$|\Phi^{k+1}\rangle = |\Phi^k\rangle - (\mathbf{H}^k)^{-1} \mathbf{g}^k \quad (35)$$

where $|\Phi\rangle = \{\zeta_i\}$ is a vector containing all the exponents of the fitted basis set functions. The superscripts k and $k+1$ represent the points before and after every iteration, and \mathbf{g} and \mathbf{H} are the gradient vector and the Hessian matrix in the step k , respectively.

At every iteration the gradient vector elements are calculated in the following manner:

$$g_i = \frac{\partial \mathcal{E}^{(2)}}{\partial \zeta_i} = x_i^4 \frac{\partial Z_{ii}}{\partial \zeta_i} + x_i^2 \sum_{j \neq i} x_j^2 \frac{\partial Z_{ij}}{\partial \zeta_i} - 2x_i^2 \frac{\partial b_i}{\partial \zeta_i} \quad (36)$$

and the expression of the Hessian matrix elements are given by:

$$\begin{aligned}
 h_{ii} &= \frac{\partial^2 \mathcal{E}^{(2)}}{\partial \zeta_i^2} = x_i^4 \frac{\partial^2 Z_{ii}}{\partial \zeta_i^2} + x_i^2 \sum_{j \neq i} x_j^2 \frac{\partial^2 Z_{ij}}{\partial \zeta_i^2} \\
 &\quad - 2x_i^2 \frac{\partial^2 b_i}{\partial \zeta_i^2} \\
 h_{ij} &= \frac{\partial^2 \mathcal{E}^{(2)}}{\partial \zeta_i \partial \zeta_j} = x_i^2 x_j^2 \frac{\partial^2 Z_{ij}}{\partial \zeta_i \partial \zeta_j}
 \end{aligned} \quad (37)$$

The Newton method is generally reliable and reasonably efficient. The optimization stops when the difference of the quadratic error integral function between two successive steps is less than a given threshold.

FIRST AND SECOND DERIVATIVES OF QUADRATIC ERROR INTEGRAL FUNCTION

The gradient vector and Hessian matrix, corresponding to eq. (36) and (37), should be computed to solve the Newton eq. (35). These expressions require the following derivatives:

$$\frac{\partial Z_{ii}}{\partial \zeta_i}, \frac{\partial Z_{ij}}{\partial \zeta_i}, \frac{\partial b_i}{\partial \zeta_i} \quad (38)$$

and:

$$\frac{\partial^2 Z_{ii}}{\partial \zeta_i^2}, \frac{\partial^2 Z_{ij}}{\partial \zeta_i^2}, \frac{\partial^2 b_i}{\partial \zeta_i^2}, \frac{\partial^2 Z_{ij}}{\partial \zeta_i \partial \zeta_j} \quad (39)$$

A normalized nS Gaussian function in the spherical polar coordinate system may be written, after integrating the angular part, as:

$$S_i(r) = N_i r^{n_i-1} e^{-\zeta_i r^2} \wedge N_i = \left[\frac{2^{2n_i-1/2} \zeta_i^{n_i+1/2}}{(2n_i-1)!! \pi^{3/2}} \right]^{1/2} \quad (40)$$

where n_i is the principal quantum number of the i th atomic shell and N_i the normalization constant. A factor $(1/4\pi)^{1/2}$ is included in the normalization factor, coming from the integration of the angular part.

Substituting expression (40) into eq. (14), the Z_{ij} measure becomes:

$$Z_{ij} = 4\pi N_i^2 N_j^2 \int_0^\infty r^{2(n_i+n_j-1)} e^{-2(\zeta_i+\zeta_j)r^2} dr \quad (41)$$

and integrating this equation yields:

$$Z_{ij} = \left(\frac{2}{\pi} \right)^{3/2} \left\{ \frac{[2(n_i+n_j)-3]!!}{(2n_i-1)!!(2n_j-1)!!} \right\} \times \zeta_i^{n_i+1/2} \zeta_j^{n_j+1/2} (\zeta_i + \zeta_j)^{1/2-n_i-n_j} \quad (42)$$

In the case where $i = j$ eq. (42) is simplified to:

$$Z_{ii} = \left\{ \frac{2^{2-2n_i}(4n_i-3)!!}{\pi^{3/2}[(2n_i-1)!!]^2} \right\} \zeta_i^{3/2} \quad (43)$$

Differentiating Z_{ij} with respect to the exponent ζ_i gives:

$$\frac{\partial Z_{ij}}{\partial \zeta_i} = Z_{ij} \left(\frac{n_i+1/2}{\zeta_i} - \frac{n_i+n_j-1/2}{\zeta_i+\zeta_j} \right) \quad (44)$$

and, in the same way, Z_{ii} measure produces:

$$\frac{\partial Z_{ii}}{\partial \zeta_i} = Z_{ii} \frac{3}{2\zeta_i} \quad (45)$$

To evaluate the \mathbf{b} vector elements, the spherical functions, $S_i(r)$, of expression (40) are substituted into eq. (18):

$$b_i = 4\pi N_i^2 \sum_{\mu,\nu} D_{\mu\nu} N_\mu N_\nu \times \int_0^\infty r^{2n_i+n_\mu+n_\nu-2} e^{-(2\zeta_i+\zeta_\mu+\zeta_\nu)r^2} dr \quad (46)$$

These integrals depend on the exponent of r . Defining $m = 2n_i + n_\mu + n_\nu - 2$, then b_i is equal to:

$$b_i = \sum_{\mu,\nu} D_{\mu\nu} N_\mu N_\nu \times \left[\frac{(2n_i+n_\mu+n_\nu-3)!! 2^{\frac{2n_i-n_\mu-n_\nu+3}{2}}}{(2n_i-1)!!} \right] \times \frac{(\zeta_i)^{n_i+1/2}}{(2\zeta_i+\zeta_\mu+\zeta_\nu)^{\frac{2n_i+n_\mu+n_\nu-1}{2}}} \quad (47)$$

if m is an even number, and if it is odd:

$$b_i = \sum_{\mu,\nu} D_{\mu\nu} N_\mu N_\nu \times \left\{ \frac{[1/2(2n_i+n_\mu+n_\nu-3)]! 2^{2n_i+1/2}}{(2n_i-1)!!} \right\} \times \frac{(\zeta_i)^{n_i+1/2}}{(2\zeta_i+\zeta_\mu+\zeta_\nu)^{\frac{2n_i+n_\mu+n_\nu-1}{2}}} \quad (48)$$

In both solutions, the b_i integral may be written using an auxiliary function B , which depends on the i th atomic shell and the variable m :

$$b_i = \sum_{\mu,\nu} B_{\mu\nu}(i, m) \quad (49)$$

Using this notation, the derivative of b_i with respect to the ζ_i exponent, is given by:

$$\frac{\partial b_i}{\partial \zeta_i} = \sum_{\mu,\nu} \left(\frac{n_i+1/2}{\zeta_i} - \frac{2n_i+n_\mu+n_\nu-1}{2\zeta_i+\zeta_\mu+\zeta_\nu} \right) \times B_{\mu\nu}(i, m) \quad (50)$$

Inserting eq. (44), (45), and (50) into eq. (36), the following expression can be obtained:

$$g_i = x_i^4 Z_{ii} \frac{3}{2\zeta_i} + x_i^2 \sum_{j \neq i} x_j^2 Z_{ij} \times \left(\frac{n_i + 1/2}{\zeta_i} - \frac{n_i + n_j - 1/2}{\zeta_i + \zeta_j} \right) - 2x_i^2 \left[\sum_{\mu, \nu} \left(\frac{n_i + 1/2}{\zeta_i} - \frac{2n_i + n_\mu + n_\nu - 1}{2\zeta_i + \zeta_\mu + \zeta_\nu} \right) \times B_{\mu\nu}(i, m) \right] \quad (51)$$

The second derivatives needed to calculate the h_{ii} Hessian matrix elements are:

$$\frac{\partial^2 Z_{ii}}{\partial \zeta_i^2} = Z_{ii} \frac{3}{4\zeta_i^2} \quad (52)$$

and:

$$\frac{\partial^2 Z_{ij}}{\partial \zeta_i^2} = Z_{ij} \left[\left(\frac{n_i + 1/2}{\zeta_i} - \frac{n_i + n_j - 1/2}{\zeta_i + \zeta_j} \right)^2 - \frac{n_i + 1/2}{\zeta_i^2} + \frac{n_i + n_j - 1/2}{(\zeta_i + \zeta_j)^2} \right] \quad (53)$$

Second derivatives, with respect to the b_i integral, are obtained as:

$$\frac{\partial^2 b_i}{\partial \zeta_i^2} = \sum_{\mu, \nu} \left[\left(\frac{n_i + 1/2}{\zeta_i} - \frac{2n_i + n_\mu + n_\nu - 1}{2\zeta_i + \zeta_\mu + \zeta_\nu} \right)^2 - \frac{n_i + 1/2}{\zeta_i^2} + \frac{2n_i + n_\mu + n_\nu - 1}{(2\zeta_i + \zeta_\mu + \zeta_\nu)^2} \right] \times B_{\mu\nu}(i, m) \quad (54)$$

Substitution of these results into eq. (37) yields the Hessian matrix elements; for the diagonal part:

$$h_{ii} = x_i^4 Z_{ii} \frac{3}{4\zeta_i^2} + x_i^2 \sum_{j \neq i} x_j^2 Z_{ij} \times \left[\left(\frac{n_i + 1/2}{\zeta_i} - \frac{n_i + n_j - 1/2}{\zeta_i + \zeta_j} \right)^2 - \frac{n_i + 1/2}{\zeta_i^2} + \frac{n_i + n_j - 1/2}{(\zeta_i + \zeta_j)^2} \right]$$

$$- 2x_i^2 \sum_{\mu, \nu} \left[\left(\frac{n_i + 1/2}{\zeta_i} - \frac{2n_i + n_\mu + n_\nu - 1}{2\zeta_i + \zeta_\mu + \zeta_\nu} \right)^2 - \frac{n_i + 1/2}{\zeta_i^2} + \frac{2n_i + n_\mu + n_\nu - 1}{(2\zeta_i + \zeta_\mu + \zeta_\nu)^2} \right] B_{\mu\nu}(i, m) \quad (55)$$

whereas the expression for the off-diagonal elements is:

$$h_{ij} = \frac{\partial^2 Z_{ij}}{\partial \zeta_i \partial \zeta_j} = Z_{ij} \left[\left(\frac{n_i + 1/2}{\zeta_i} - \frac{n_i + n_j - 1/2}{\zeta_i + \zeta_j} \right) \times \left(\frac{n_j + 1/2}{\zeta_j} - \frac{n_i + n_j - 1/2}{\zeta_i + \zeta_j} \right) + \frac{n_i + n_j - 1/2}{(\zeta_i + \zeta_j)^2} \right] \quad (56)$$

Computational Scheme

A program called GATOMIC²² has been constructed to compute fitted atomic shells using *nS-type* Gaussian functions. Figure 1 describes the main steps to obtain the coefficients and exponents of the ASA functions.

The initial ζ_i exponents are taken from an *even-tempered* geometric sequence:²³

$$\zeta_i = \frac{1}{n_i} \alpha \beta^i, i = 1, 2, \dots, N \quad (57)$$

where n_i is the principal quantum number of the i th spherical shell, and N describes the number of used fitting functions. To perform a global search of the exponents, a grid method is employed to explore the surface described by the *even-tempered* set generating the parameters $\{\alpha, \beta\}$. Then, the computational procedure described in Figure 1 is carried out over every point of this grid.

As previously indicated, in the first optimization step, the similarity matrix, \mathbf{Z} , is diagonalized and the $\{x_i\}$ coefficient set is extracted from the eigenvector producing the minimal $\varepsilon^{(2)}$ value. Then, using EJR, as described previously, ASA


```

DATA: N = NUMBER OF FITTING FUNCTIONS
a) COMPUTE  $\zeta_i$  EXPONENTS FROM EVEN-TEMPERED SEQUENCE
   CALCULATE  $Z_{ab}$ ,  $b$  AND  $Z$ 
b) DIAGONALIZE THE MATRIX  $Z$  AND OBTAIN THE EIGENVALUES AND EIGENVECTORS ( $v$ )
   DO FOR  $j=1$  TO  $N$ 
        $x_i = v_{ij} \wedge w_i = x_i^2, i=1, \dots, N$ 
        $\varepsilon^{(2)} = \min(\varepsilon^{(2)}, Z_{ab} + w^T Z w - 2b^T w)$ 
   END DO FOR  $j$ 
    $\varepsilon^{(2)} = \varepsilon^{(2)}$ 
c) COEFFICIENTS OPTIMIZATION USING EJR
   DO WHILE (.true.)
       DO FOR  $p=1$  TO  $N-1$ 
           DO FOR  $q=p+1$  TO  $N$ 
               COMPUTE  $E_{04}$ ,  $E_{13}$ ,  $E_{22}$  AND  $E_{11}$  FROM EQUATION (33)
               GRID SEARCH TO FIND INITIAL SQUARE SINE:  $s^2$ 
               DO WHILE (.true.)
                   COMPUTE  $T_1$ ,  $T_2$  AND  $T_3$  FROM EQUATION (34)
                    $t_1 = \frac{T_2 - \sqrt{T_2^2 + T_1 T_3}}{T_1} \rightarrow s_1, c_1 \wedge \delta \varepsilon^{(2)} = E_{04} s_1^4 + E_{13} c_1 s_1^2 + E_{22} s_1^2 + E_{11} c_1 s_1$ 
                    $t_2 = \frac{T_2 - \sqrt{T_2^2 + T_1 T_3}}{T_1} \rightarrow s_2, c_2 \wedge \delta \varepsilon^{(2)} = E_{22} s_2^4 + E_{13} c_2 s_2^2 + E_{04} s_2^2 + E_{11} c_2 s_2$ 
                    $\delta \varepsilon^{(2)} = \min(\delta \varepsilon^{(2)}, \delta \varepsilon^{(2)}) \rightarrow s^2$ 
                   IF  $|s^2 - s^2| < \text{TOLER}_2$  THEN
                        $x_p \leftarrow c x_p - s x_q$ 
                        $x_q \leftarrow s x_p + c x_q$ 
                        $w_p = x_p^2 \wedge w_q = x_q^2$ 
                   EXIT DO WHILE_2
               END IF
                $s^2 = s^2$ 
           END DO WHILE_2
       END DO FOR  $q$ 
   END DO FOR  $p$ 
    $\varepsilon^{(2)} = Z_{ab} + w^T Z w - 2b^T w$ 
   IF  $|\varepsilon^{(2)} - \varepsilon^{(2)}| < \text{TOLER}_1$  EXIT DO WHILE_1
    $\varepsilon^{(2)} = \varepsilon^{(2)}$ 
END DO WHILE_1
d) EXPONENTS OPTIMIZATION
DO WHILE (.true.)
    CALL NEWTON SEARCH: NEEDS THE COMPUTATION OF  $g$  AND  $H$  (EQUATION (35))
    CALCULATE  $b$  AND  $Z$ 
    CALL EJR (SECCION (c))
     $\varepsilon^{(2)} = Z_{ab} + w^T Z w - 2b^T w$ 
    IF  $|\varepsilon^{(2)} - \varepsilon^{(2)}| < \text{TOLER}$  EXIT DO WHILE
     $\varepsilon^{(2)} = \varepsilon^{(2)}$ 
END DO WHILE
    
```

FIGURE 1. Computational flowchart describing the optimization of ASA coefficients and exponents.

coefficients are optimized solving the second degree equation, appearing in expression (34), using an iterative procedure (section c in Fig. 1). Next, the optimization of ζ_i exponents begins. After the Newton algorithm a new EJR process is again carried out to obtain the most accurate fitted positive coefficients, and an iterative process is repeated until the convergence criterion is fulfilled.

To produce an example to depict the variation of $\{\alpha, \beta\}$ even-tempered parameters, Figure 2 shows the surface obtained for the chlorine atom when five 1S Gaussian functions are used ($N = 5$). A 30-point grid for each parameter α and β is per-

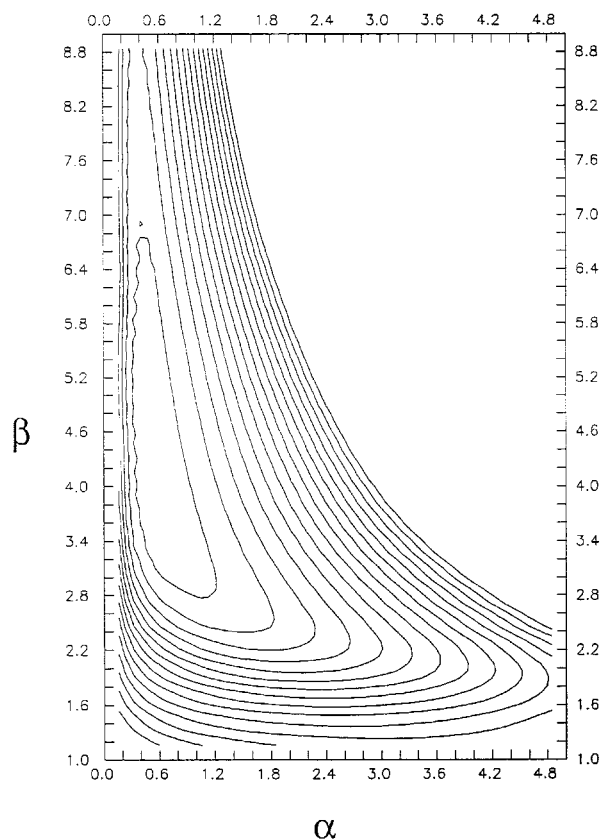


FIGURE 2. Surface described by the variation of α and β even-tempered parameters for the chlorine atom.

formed to explore the surface in the interval $\alpha \in [0.01, 5.0]$ and $\beta \in [1.0, 9.0]$. Over every point of the grid, EJR optimization of the $\{x_i\}$ coefficients is performed, keeping the ASA exponents invariable. The minimum value of $\varepsilon^{(2)}$ is 0.0143 a.u. and has been obtained at $\alpha = 0.1734$ and $\beta = 5.1379$ parameter values. As will be seen in the next section, this $\varepsilon^{(2)}$ value may be improved if exponents are optimized and the exponent even-tempered sequence abandoned.

Results

Two types of results will be presented. First, to evaluate the fitting procedure, atomic QS-SM are computed using ASA functions and are compared with the *ab initio* ones. A following analysis will verify the behavior of the ASA atomic basis set density functions in molecular calculations, using the *promolecular* approximation, as described in eq. (11).

ATOMIC CALCULATIONS

Table I lists *ab initio* QS-SM (Z_{aa}) using the 3-21G basis set²⁴ for atoms Li to Kr. HF *ab initio* atomic density functions have been calculated with the ATOMIC-95 program.²⁵ Any GTO basis set could have been chosen. The present calculation uses the 3-12G level because it is available in the Gaussian-94 program²⁶ from H to Kr. Systematic calculations on other basis sets are under study.

Different 1S Gaussian basis sets have been computed with the GATOMIC program, which uses as the input data the output data of the ATOMIC

program. Table I also gives the quadratic error integral function and the relative error produced in the QS-SM, using different numbers of functions per atom. Table I does not show results for heavy atoms using three and four functions (corresponding to columns $N = 3$ and $N = 4$), because it is not possible to achieve a good fitting with fewer functions. Hydrogen and helium atoms are not included in this table because their contributions to MQSM are negligible (see the QSM map HCCH—Ne in the next subsection), and only one function is used to describe these two atoms. Coeffi-

TABLE I.
Results Using a 3-21G Basis Set²⁴ for Li to Kr.^a

Atom (electronic state)	Z_{aa} (a.u.)	N	$\varepsilon^{(2)}$	% Z_{aa} error	N	$\varepsilon^{(2)}$	% Z_{aa} error	N	$\varepsilon^{(2)}$	% Z_{aa} error
Li(² S)	0.34440	5	0.0000023	−0.0218	4	0.000016	−0.0111	3	0.00085	0.0316
Be(¹ S)	0.51914	5	0.0000064	−0.0330	4	0.000009	−0.0095	3	0.00111	0.0775
B(² P)	0.69161	5	0.0000075	0.0137	4	0.000027	−0.0641	3	0.00141	0.0035
C(³ P)	0.87083	5	0.0000174	−0.0021	4	0.000074	−0.1214	3	0.00181	−0.0904
N(⁴ S)	1.05761	5	0.0000179	−0.0023	4	0.000164	−0.1938	3	0.00239	−0.1991
O(³ P)	1.25220	5	0.0000195	−0.0179	4	0.000345	−0.2861	3	0.00316	−0.3427
F(² P)	1.45639	5	0.0000185	0.0079	4	0.000602	−0.3855	3	0.00413	−0.5049
Ne(¹ S)	1.67148	5	0.0000245	0.0133	4	0.001046	−0.5552	3	0.00532	−0.6790
Na(² S)	1.90067	5	0.0001318	−0.0922	4	0.002114	−0.3195	3	0.00724	−1.3363
Mg(¹ S)	2.14464	5	0.0002249	−0.0351	4	0.002733	−0.0990	3	0.01172	−2.0284
Al(² P)	2.38804	5	0.0002282	−0.0154	4	0.003055	0.0347	3	0.01888	−2.6856
Si(³ P)	2.63649	5	0.0002436	−0.0038	4	0.003159	0.0641	3	0.02900	−3.2867
P(⁴ S)	2.88964	5	0.0001657	−0.0128	4	0.003197	0.0672	3	0.04829	−2.8037
S(³ P)	3.14729	5	0.0001382	−0.0148	4	0.003237	0.0659	3	0.07193	1.9657
Cl(² P)	3.40948	5	0.0001097	−0.0167	4	0.003261	0.0566	3	0.07578	1.9987
Ar(¹ S)	3.67627	5	0.0000906	−0.0186	4	0.003282	0.0516	3	0.07941	1.9736
K(² S)	3.94005	5	0.0001216	−0.0430						
Ca(¹ S)	4.21566	5	0.0002768	−0.0808						
Sc(² D)	4.48734	5	0.0003525	−0.0933						
Ti(³ F)	4.76039	5	0.0004101	−0.1030						
V(⁴ F)	5.03508	5	0.0004908	−0.1182						
Cr(⁵ D)	5.30329	5	0.0005757	−0.1316						
Mn(⁶ S)	5.59114	5	0.0006672	−0.1438						
Fe(⁵ D)	5.87170	5	0.0007898	−0.1604						
Co(⁴ F)	6.15446	5	0.0009223	−0.1717						
Ni(³ F)	6.44012	5	0.0010698	−0.1948						
Cu(² D)	6.72772	5	0.0012540	−0.2118						
Zn(¹ S)	7.01766	5	0.0014076	−0.2365						
Ga(² P)	7.31617	5	0.0022071	−0.2992						
Ge(³ P)	7.61659	5	0.0034518	−0.3805						
As(⁴ S)	7.97677	5	0.0043724	−0.4565						
Se(³ P)	8.22790	5	0.0060097	−0.5303						
Br(² P)	8.55327	5	0.0080829	−0.6061						
Kr(¹ S)	8.84739	5	0.0106579	−0.6765						

^a*Ab initio* QS-SM (Z_{aa}), quadratic error integral function and relative error in QS-SM using different numbers (N) of 1S Gaussian functions per atom.

cients and exponents for all these basis sets of ASA functions are available for downloading at a WWW site.²⁷

SOME MOLECULAR EXAMPLES

QSM maps²⁸ constitute a very useful tool for observing the behavior of the ASA functions with respect to the *ab initio* ones.^{11b} A QSM map between a molecule, *A*, and an atom, *b*, can be defined as the integral:

$$Z_{Ab}[\mathbf{R}] = \iint \rho_A(\mathbf{r}_1) \Omega(\mathbf{r}_1, \mathbf{r}_2) \rho_b(\mathbf{r}_2 - \mathbf{R}) d\mathbf{r}_1 d\mathbf{r}_2 \quad (58)$$

where it is made explicit that the measure $Z_{Ab}[\mathbf{R}]$ depends on the atom position \mathbf{R} . Two examples are presented of QSM maps using overlap-like measures; both are obtained when a neon atom is moved along the molecular axis defined by the

atoms of a linear molecule. Density functions and molecular geometries are obtained using the Gaussian-94 program and the HF/3-21G basis set. In the first example, a resulting QSM map for the acetylene molecule is shown. In Figure 3, the solid line shows the *ab initio* QSM map, whereas the dashed line corresponds to ASA computations, using one function for H and three functions for C and Ne atoms (corresponding to column $N = 3$ of Table I). Maximal error between *ab initio* and ASA measures is -0.006 a.u. and is located in the C—C bonding region. The second example, presented in Figure 4, consists of a QSM map of nitrous oxide, NNO. In the same manner as the above example, *ab initio* calculations are drawn in solid line and ASA measures, using three fitted functions for N, O, and Ne atoms (dashed). Bonding regions N—N and N—O are the areas where ASA functions produce maximal error with a maximum value of -0.016 a.u. This effect is due to the *promolecular* approximation itself, which obviously produces a

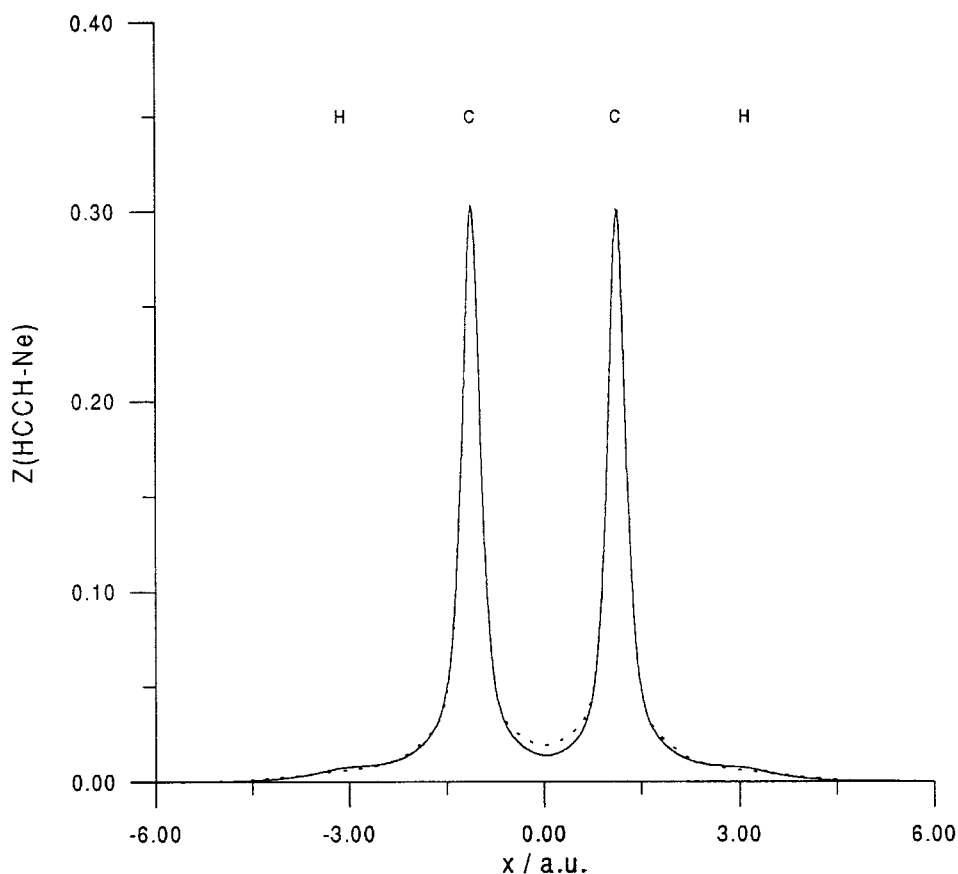


FIGURE 3. QSM map for the system acetylene–neon. Solid line corresponds to *ab initio* HF/3-21G calculation. Dashed line is the ASA calculation.

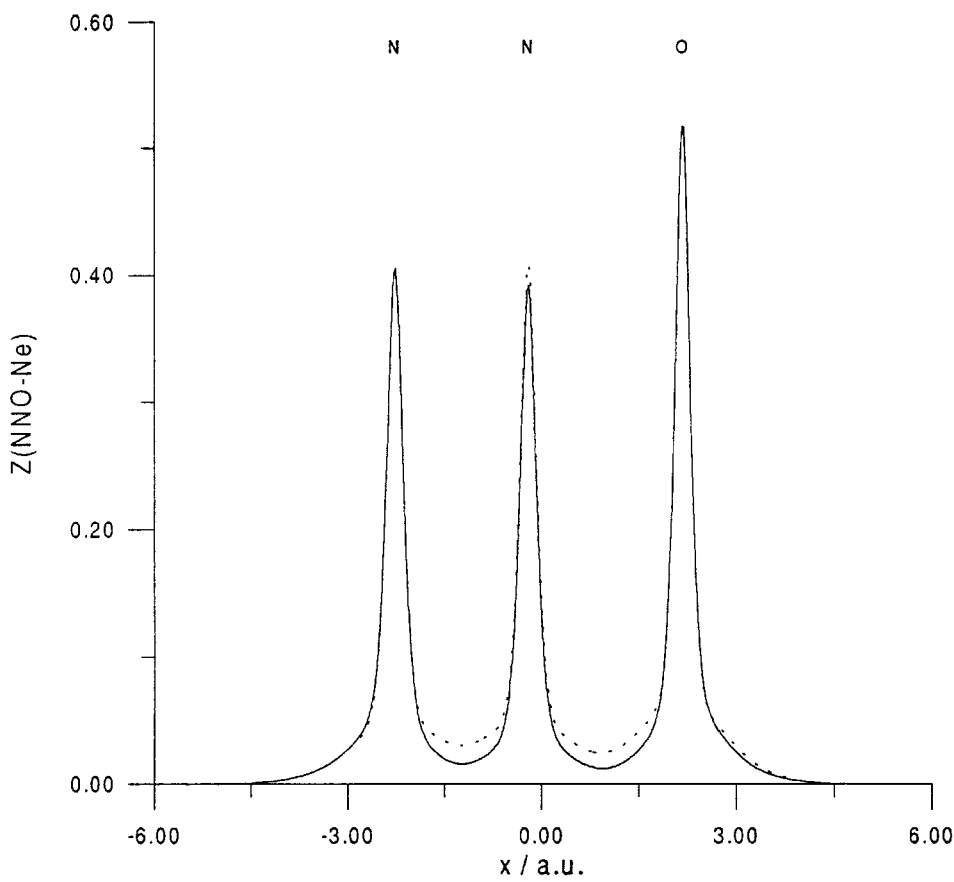


FIGURE 4. QSM map for the system nitrous oxide–neon. Solid line corresponds to *ab initio* HF / 3-21G calculation. Dashed line is the ASA calculation.

less accurate density representation in the bonding regions of molecules, but correctly describes atom nuclei. As is evidenced in the following examples, bonding regions do not have so great an influence in MQSM, and ASA results may be considered as possessing sufficient accuracy for the practical purposes of QSM.

The optimized ASA functions have been ap-

plied to a molecular set consisting of nine fluoro- and chlorosubstituted methanes. Tables II and III show *ab initio* MQSM and Carbó indices^{14a} for all pairs of molecules. Optimized geometries have been obtained with the Gaussian-94 program using the HF/3-21G 6D 10F level of theory, and optimal molecular superpositions have been found using the method described in ref. 18. This molec-

TABLE II.
Ab initio MQSM for Fluoro- and Chlorosubstituted Methanes Using HF / 3-21G Densities.

	CH ₄	CFH ₃	CF ₂ H ₂	CF ₃ H	CF ₄	CClH ₃	CCl ₂ H ₂	CCl ₃ H	CCl ₄
CH ₄	0.31481	0.32213	0.22321	0.17084	0.13841	0.54534	0.33762	0.24449	0.19162
CFH ₃		0.46106	0.31639	0.23657	0.18794	0.66778	0.41364	0.29967	0.23495
CF ₂ H ₂			0.39548	0.27592	0.20845	0.46638	0.28836	0.20889	0.16388
CF ₃ H				0.33334	0.26499	0.35777	0.22114	0.16046	0.12593
CF ₄					0.28533	0.29006	0.17914	0.13001	0.10246
CClH ₃						1.50383	0.92818	0.67035	0.52526
CCl ₂ H ₂							1.13480	0.71621	0.45926
CCl ₃ H								0.88792	0.63259
CCl ₄									0.72536

TABLE III.

Ab initio Carbó Indices for Fluoro- and Chlorosubstituted Methanes Using HF / 3-21G Densities.

	CH ₄	CFH ₃	CF ₂ H ₂	CF ₃ H	CF ₄	CClH ₃	CCl ₂ H ₂	CCl ₃ H	CCl ₄
CH ₄	1	0.84553	0.63260	0.52739	0.46181	0.79258	0.56486	0.46243	0.40100
CFH ₃		1	0.74094	0.60345	0.51815	0.80196	0.57185	0.46836	0.40628
CF ₂ H ₂			1	0.75994	0.62054	0.60476	0.43044	0.35251	0.30597
CF ₃ H				1	0.85926	0.50531	0.35956	0.29494	0.25610
CF ₄					1	0.44281	0.31482	0.25829	0.22522
CClH ₃						1	0.71051	0.58011	0.50292
CCl ₂ H ₂							1	0.71350	0.50620
CCl ₃ H								1	0.78824
CCl ₄									1

ular set has been used in previous studies.^{11, 14f} as a benchmark test to verify new fitting algorithms and approximations.

Table IV resumes the results obtained using different sets of ASA functions described in Table I. Main error scores for the molecular set studied are presented in Table IV. The data provided in the table consist of the following parameters: greatest error, error arithmetic mean, and standard deviation produced when relative error for all pairs of molecules is analyzed. As was expected, the accuracy achieved increases when the number of ASA functions per atom are augmented. Another aspect that arises from Table IV corresponds to the fact that ASA Carbó indices agree with *ab initio* values, in a better way than MQSM. This characteristic has already been observed in previous studies,^{11, 14f} and may be an effect caused by the normalization of the measure, performed when Carbó indices are computed.

The present results are the best obtained in our laboratory using a *promolecular* approximation up to now. This is shown when they are compared

with those computed in ref. 11b, where 1S function exponents were fitted to reproduce atomic QS-SM. Another evaluation may be obtained when the present results of *promolecular* ASA are compared with those of molecular ASA.^{11a} In ref. 11a, the set of nine fluoro- and chlorosubstituted methanes have also been studied, although at a different level of theory (HF/6-31G*). This last remark does not become an inconvenience at all, if ref. 11a error and standard deviation means are compared with those of Table IV. In the molecular ASA computations found in ref. 11a, the percent error arithmetic mean between *ab initio* and fitted MQSM is 0.056%, with a standard deviation of 0.05%. In contrast, within Carbó index calculations of the same source, the relative error arithmetic mean is 0.066% and the standard deviation becomes 0.06%. The accuracy characteristics of these quoted results are comparable with the five-function-level *promolecular* measures of the present work (see Table IV) and the effect is more pronounced within the Carbó index figures.

To show an example of the ASA results, in

TABLE IV.

Relative Errors with Associated Greatest Error Value, Arithmetic Mean and Standard Deviation Produced by Comparing *Ab Initio* Values (Tables II and III) with ASA MQSM and Carbó Indices for Fluoro- and Chlorosubstituted Methanes.

Number of shells				Measure	Relative errors			
H	C	F	Cl		Greatest error (%) / pair of molecules		Mean of errors	Standard deviation
1	3	3	4	Z _{AB}	-1.421	(CH ₄ — CH ₄)	0.534	0.380
				C _{AB}	1.282	(CFH ₃ — CClH ₃)	0.576	0.537
1	4	4	4	Z _{AB}	-1.432	(CH ₄ — CH ₄)	0.350	0.355
				C _{AB}	0.521	(CFH ₃ — CClH ₃)	0.173	0.189
1	5	5	5	Z _{AB}	-1.277	(CH ₄ — CH ₄)	0.294	0.320
				C _{AB}	-0.244	(CH ₄ — CCl ₄)	0.052	0.071

TABLE V.
MQSM for Fluor- and Chlorosubstituted Methanes Using Five-Function Level ASA Fitted Densities.

	CH ₄	CFH ₃	CF ₂ H ₂	CF ₃ H	CF ₄	CClH ₃	CCl ₂ H ₂	CCl ₃ H	CCl ₄
CH ₄	0.31883	0.32487	0.22503	0.17217	0.13945	0.55023	0.34065	0.24669	0.19335
CFH ₃		0.46273	0.31742	0.23708	0.18803	0.66842	0.41406	0.30002	0.23527
CF ₂ H ₂			0.39648	0.27668	0.20896	0.46703	0.28873	0.20916	0.16412
CF ₃ H				0.33405	0.26549	0.35822	0.22139	0.16070	0.12613
CF ₄					0.28590	0.29041	0.17934	0.13019	0.10265
CClH ₃						1.50494	0.92893	0.67097	0.52583
CCl ₂ H ₂							1.13538	0.71670	0.45975
CCl ₃ H								0.88831	0.63296
CCl ₄									0.72567

Tables V and VI, MQSM and Carbó indices of the set of fluoro- and chlorosubstituted methanes are presented in five-function-level calculation. To test the behavior of the present ASA functions in calculations involving large molecules with heavy atoms, a sample of eight molecules has been chosen. These molecular structures have been taken from the Cambridge Structural Database (CSD)²⁹ and are shown in Table VII.

Table VIII lists the results of molecular QS-SM using *ab initio* and *promolecular* ASA density functions. From the CSD geometries, molecular electronic density functions have been calculated with the Gaussian-94 program using the HF/3-21G 6D 10F level. ASA QS-SM were computed from the atomic basis presented in Table I. The rule was to use one function for the hydrogen atom, three functions for all atoms belonging to the second period, four functions for the atoms of the third period, and five functions for the atoms of the fourth period. This procedure gives the best results with the minimum number of atomic ASA functions.

The results given in Table VIII can be considered satisfactory, in view of the fact that the relative errors made in fitted QS-SM are less than 0.81% for all the molecules studied. *Promolecular* functions produce an electron density excess around the zone of the bonding regions, as is shown in Figures 3 and 4. This less accurate density representation can also be evidenced by noticing that fitted QS-SM in Table VIII are *always* greater than the *ab initio* exact values.

Conclusions

The most important idea to be extracted from the present results, is that the EJR technique can be a useful methodological procedure that allows us to obtain constrained fitted density functions with positive coefficients, and in this manner preserving the statistical meaning of the fitted density functions. Also, the optimization of Gaussian scale factors, using a Newton method, gives even more accurate ASA functions. Within these techniques,

TABLE VI.
Carbó Indices for Fluoro- and Chlorosubstituted Methanes Using Five-Function Level ASA Fitted Densities.

	CH ₄	CFH ₃	CF ₂ H ₂	CF ₃ H	CF ₄	CClH ₃	CCl ₂ H ₂	CCl ₃ H	CCl ₄
CH ₄	1	0.84579	0.63291	0.52758	0.46188	0.79434	0.56619	0.46355	0.40198
CFH ₃		1	0.74107	0.60301	0.51696	0.80098	0.57125	0.46795	0.40600
CF ₂ H ₂			1	0.76025	0.62064	0.60462	0.43034	0.35245	0.30596
CF ₃ H				1	0.85909	0.50523	0.35949	0.29500	0.25619
CF ₄					1	0.44274	0.31477	0.25835	0.22537
CClH ₃						1	0.71064	0.58031	0.50317
CCl ₂ H ₂							1	0.71364	0.50651
CCl ₃ H								1	0.78836
CCl ₄									1

TABLE VII.
Cambridge Structural Database Reference Codes, Chemical Formulae, and Structures of Selected Molecules
Used to Compute QS-SM.

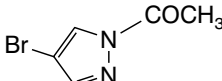
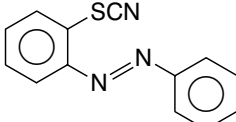
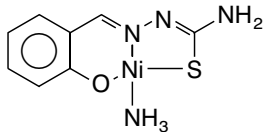
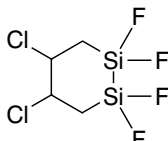
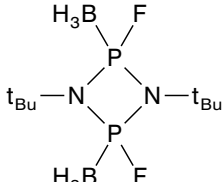
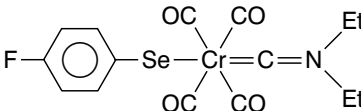
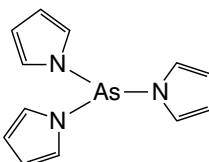
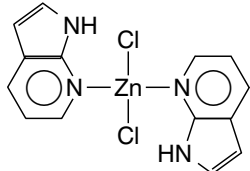
Molecule	CSD ref. code ²⁹	Formula	Structure
A ₁	ABPZOL10	C ₅ H ₅ BrN ₂ O	
A ₂	ABSFCN	C ₁₃ H ₉ N ₃ S	
A ₃	AMSCNI11	C ₈ H ₁₀ N ₄ NiOS	
A ₄	BABPAV	C ₄ H ₆ Cl ₂ F ₄ Si ₂	
A ₅	BAKNAC	C ₈ H ₂₄ B ₂ F ₂ N ₂ P ₂	
A ₆	BAMCOH	C ₁₅ H ₁₄ CrFNO ₄ Se	
A ₇	BEYCIR	C ₁₂ H ₁₂ AsN ₃	
A ₈	BIFXAP	C ₁₄ H ₁₂ Cl ₂ N ₄ Zn	

TABLE VIII.
QS-SM (Z_{AA}) Using HF/3-21G and ASA Densities
and Relative Error for the Molecular Set in Table VII.

	Z_{AA}^{HF} (a.u.)	N^a	Z_{AA}^{ASA} (a.u.)	% Z_{AA}
A ₁	1.27877	34	1.28601	−0.5659
A ₂	0.08890	61	0.08932	−0.4733
A ₃	0.33386	58	0.33646	−0.7776
A ₄	0.21973	46	0.22001	−0.1278
A ₅	0.09287	74	0.09317	−0.3287
A ₆	0.30662	87	0.30817	−0.5040
A ₇	0.48412	62	0.48642	−0.4734
A ₈	0.25129	79	0.25331	−0.8025

^a N is the number of shells per molecule (see text for number of shells per atom).

function exponents and coefficients are completely optimized. In this context it has been demonstrated that an excellent ASA Gaussian spherical basis set can be obtained.

Moreover, the present work shows how *promolecular* electronic densities describe molecular densities with sufficient accuracy for some MQSM simplified computational purposes. In subsequent MQSM studies, it will only be necessary to know the molecular coordinates, and then, with a set of ASA functions, as parameterized in this work, the molecular electronic density distribution can be built automatically without further effort.

Acknowledgments

The authors thank the referees for their advice on improving the contents and structure of the present work.

References

- (a) K. J. Miller and K. Ruedenberg, *J. Chem. Phys.*, **48**, 3414 (1968); (b) R. C. Raffanetti and K. Ruedenberg, *Int. J. Quant. Chem.*, **35**, 625 (1970); (c) D. K. Hoffman, R. C. Raffanetti and K. Ruedenberg, *J. Math. Phys.*, **13**, 528 (1972).
- (a) C. Edmiston and K. Ruedenberg, *Rev. Mod. Phys.*, **35**, 457 (1963); (b) R. C. Raffanetti, K. Ruedenberg, C. L. Janssen and H. F. Schaefer, *Theor. Chim. Acta*, **86**, 149 (1993).
- D. M. Silver, E. L. Mehler and K. Ruedenberg, *J. Chem. Phys.*, **52**, 1206 (1970).
- (a) R. Carbó, L. Domingo and J. J. Peris, *Adv. in Quant. Chem.*, **15**, 215 (1982); (b) R. Carbó, J. Miró, L. Domingo and J. J. Novoa, *Adv. in Quant. Chem.*, **20**, 375 (1989); (c) R. Carbó, L. Domingo, J. J. Peris and J. J. Novoa, *J. Mol. Struct.*, **93**, 15 (1983); (d) R. Carbó and B. Calabuig, *Comput. Phys. Commun.*, **52**, 345 (1989).
- R. Carbó and L. Domingo, *Int. J. Quant. Chem.*, **23**, 517 (1987).
- C. G. J. Jacobi, *J. Reine Angew. Math.*, **30**, 51 (1846).
- R. Carbó, L. Molino and B. Calabuig, *J. Comput. Chem.*, **13**, 155 (1992).
- R. Carbó-Dorca and E. Besalú, *J. Math. Chem.*, **20**, 263 (1997).
- J. Mestres, M. Solà, M. Duran and R. Carbó, *J. Comput. Chem.*, **15**, 1113 (1994).
- J. Cioslowski, P. Piskorz and P. Rez, *J. Chem. Phys.*, **106**, 3607 (1997).
- (a) P. Constans and R. Carbó, *J. Chem. Inf. Comput. Sci.*, **35**, 1046 (1995); (b) P. Constans, L. Amat, X. Fradera and R. Carbó-Dorca, In R. Carbó-Dorca and P. G. Mezey (Eds.), *Advances in Molecular Similarity*, Vol. 1, JAI Press, Inc., Greenwich, CT, 1996, p. 187.
- J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, Princeton, N.J., 1955.
- See, for example: (a) M. A. Johnson and G. Maggiora, Eds., *Concepts and Applications of Molecular Similarity*, Wiley, New York, 1990; (b) R. Carbó, Ed., *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, Kluwer Amsterdam, 1995; (c) R. Carbó-Dorca, and P. G. Mezey, Eds., *Advances in Molecular Similarity*, Vol. 1, JAI Press, Greenwich, CT, 1996.
- See, for example: (a) R. Carbó, M. Arnau and L. Leyda, *Int. J. Quantum Chem.*, **17**, 1185 (1980); (b) R. Carbó and B. Calabuig, *Comput. Phys. Commun.*, **55**, 117 (1989); (c) R. Carbó and B. Calabuig, *Int. J. Quantum Chem.*, **42**, 1681 (1992); (d) R. Carbó, E. Besalú and L. Vera, *Adv. in Quantum Chem.*, **25**, 253 (1994); (e) M. Solà, J. Mestres, R. Carbó and M. Duran, *J. Am. Chem. Soc.*, **116**, 5909 (1994); (f) E. Besalú, R. Carbó, J. Mestres and M. Solà, *Topics Curr. Chem.*, **173**, 31 (1995); (g) J. Mestres, M. Solà, R. Carbó, F. J. Luque and M. Orozco, *J. Phys. Chem.*, **100**, 606 (1996).
- See, for example: (a) D. L. Cooper and N. L. Allan, *J. Chem. Soc. Faraday Trans.*, **83**, 449 (1987); (b) D. L. Cooper and N. L. Allan, *J. Am. Chem. Soc.*, **114**, 4773 (1992); (c) J. Cioslowski and E. D. Fleischmann, *J. Am. Chem. Soc.*, **113**, 64 (1991); (d) J. V. Ortiz and J. Cioslowski, *Chem. Phys. Lett.*, **185**, 270 (1991); (e) R. Ponec and M. Strnad, *J. Phys. Org. Chem.*, **4**, 701 (1991); (f) R. Ponec and M. Strnad, *Int. J. Quant. Chem.*, **42**, 501 (1992); (g) E. E. Hodgkin and W. G. Richards, *Int. J. Quant. Chem.*, **14**, 105 (1987); (h) P. G. Mezey, *J. Comp. Chem.*, **8**, 462 (1987).
- (a) K. Ruedenberg and W. H. E. Schwarz, *J. Chem. Phys.*, **92**, 4956 (1990); (b) P. Coppens, In *International Tables for Crystallography*, Vol. B, Kluwer, Amsterdam, 1992, p. 10; (c) P. Coppens and J. Becker, In *International Tables for Crystallography*, Vol. C, Kluwer, Amsterdam, 1992, p. 628.
- X. Fradera, L. Amat, E. Besalú and R. Carbó-Dorca, *Quant. Struct.-Activ. Relat.*, **16**, 25 (1977).
- (a) P. Constans, L. Amat and R. Carbó-Dorca, *J. Comput. Chem.*, **18**, 826 (1997); (b) L. Amat, P. Constans and R. Carbó, *Sci. Gerund*, **22**, 109 (1996).
- J. A. Pople and D. L. Beveridge, *Approximate Molecular Orbital Theory*, McGraw-Hill, New York, 1970.
- See, for example: D. A. Pierre, *Optimization Theory with Applications*, Wiley, New York, 1969.
- See, for example: (a) H. Kubinyi, Ed., *3D QSAR in Drug Design: Theory Methods and Applications*, ESCOM, Leiden,

- 1993; (b) P. M. Dean, Ed. *Molecular Similarity in Drug Design*, Blackie, London, 1995; (c) A. C. Good, S. S. So and W. G. Richards, *J. Med. Chem.*, **36**, 433 (1993).
22. L. Amat and R. Carbó-Dorca, *GATOMIC Program*, Institute of Computational Chemistry, University of Girona, Girona, Spain, 1997.
 23. S. Huzinaga and M. Klobukowski, *J. Mol. Struct. (Theorchem)*, **167**, 1 (1988).
 24. (a) J. S. Binkley, J. A. Pople and W. J. Hehre, *J. Am. Chem. Soc.*, **102**, 939 (1980); (b) M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro and W. J. Hehre, *J. Am. Chem. Soc.*, **104**, 2797 (1982); (c) W. J. Pietro, M. M. Francl, W. J. Hehre, D. J. DeFrees, J. A. Pople and J. S. Binkley, *J. Am. Chem. Soc.*, **104**, 5039 (1982).
 25. *ATOMIC Program 1995*, by R. Carbó-Dorca, from *A General Program for Calculation of SCF Orbitals by the Expansion Method*, B. Roos, C. Salez, A. Veillard and E. Clementi, IBM Research/RJ518(#10901), 1968.
 26. M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. A. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, Al-Laham, M. A., V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanaykkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Comperts, R. L. Martin, D. J. Fox, H. S. Binkley, D. J. Defrees, H. Baker, J. J. P. Stewart, M. Head-Gordon, C. Gonzalez and J. A. Pople, *Gaussian-94*, (Revision A.1), Gaussian, Inc., Pittsburgh, PA, 1995.
 27. ASA coefficients and exponents can be seen and downloaded from the WWWsite: <http://iqc.udg.es/cat/similarity/ASA/funcset.html>
 28. (a) R. Carbó-Dorca, E. Besalú, L. Amat and X. Fradera, In R. Carbó-Dorca and P. G. Mezey, Eds., *Advances in Molecular Similarity*, Vol. 1, JAI Press, Greenwich, CT, 1996, p. 1; (b) L. Amat, X. Fradera and R. Carbó, *Sci. Gerund.*, **22**, 97 (1996).
 29. F. H. Allen and O. Kennard, *Chem. Design Automat. News*, **8**, 31 (1993).